

CSE-587 Data Intensive Computing

Assignment-5

Accumulo

Karthik Kiran
UBIT: kkiran

INTRODUCTION

In this project, we use Twitter API to collect latest 100 tweets of NBA teams and store in CSV files. Accumulo is used to process the Twitter data and find out the popularity of every team.

IMPLEMENTATION

Twitter API run on Python script is used to gather data from twitter. There are total 30 files gathered (one for each NBA team). Each CSV file has 100 tweets. First column is the timestamp of the tweet and the second column is the tweet itself.

The popularity of teams was found out based on number of “Win” and “Lose” in the files. This was done using Accumulo. The popularity of teams in each conference was computed. The entire computation was implemented in three Mapreduce phases using Accumulo Tables. Each phase consists of 1 Mapper and 1 Reducer phase.

Phase 1:

- Files from CSV were read into an Accumulo table. Mapper 1 read all files and formed key, value pairs. Key was Hash tag of team and values were the tweets.
- Reducer 1 computed the number of occurrences of “Win” (case insensitive) and “Lose” for each team.
- Output of Reducer was written to an Accumulo table using Mutation. Key was Team Hash Tag and value was (Number of win + “!” + Number of Lose).
- 2 users were created for the table. One for “east” and one for “west”. East teams data were written into the table with access only available to “east” user and “west” teams data available only to “west” user.

Phase 2:

- Input to this phase is the Accumulo table data that is accessible only to “east” users.
- Mapper 1 produces 2 keys. One is “Win” and “Other” is “lose”. Value includes “Number of occurrences of win + “!” + Team Hash Tag for key “win” and “Number of occurrences of lose + “!” + Team Hash Tag for key “lose”.
- Reducer 2 sorts the teams based on decreasing order of losses and decreasing order of wins and writes into an Accumulo table which is accessible only to the “east” user.

Phase 3:

- Input to this phase is the Accumulo table data that is accessible only to “west” users.
- Mapper 1 produces 2 keys. One is “Win” and “Other” is “lose”. Value includes “Number of occurrences of win + “!” + Team Hash Tag for key “win” and “Number of occurrences of lose + “!” + Team Hash Tag for key “lose”.
- Reducer 2 sorts the teams based on decreasing order of losses and decreasing order of wins and writes into an Accumulo table which is accessible only to the “west” user.

OUTPUT SCREEN SHOTS

root@acc> user west

password: west

west@acc> table westwinlose

```
west@acc westwinlose> user west
Enter password for user west: ****
west@acc westwinlose> scan
WESTLOSE westlose:20080907 [] San Antonio Spurs #GoSpursGo 7
WESTLOSE westlose:20080908 [] Portland Trailblazers #TrailBlazers 4
WESTLOSE westlose:20080909 [] LA Lakers #Lakers 3
WESTLOSE westlose:20080910 [] Golden State Warriors #GSWarriors 3
WESTLOSE westlose:20080911 [] Denver Nuggets #Nuggets 2
WESTLOSE westlose:20080912 [] Dallas Mavericks #Mavs 2
WESTLOSE westlose:20080913 [] Minnesota Timberwolves #TWolves 1
WESTLOSE westlose:20080914 [] Oklahoma City #okcthunder 0
WESTLOSE westlose:20080915 [] Utah Jazz #UtahJazz 0
WESTLOSE westlose:20080916 [] Phoenix Suns #Suns 0
WESTLOSE westlose:20080917 [] Houston Rockets #Rockets 0
WESTLOSE westlose:20080918 [] Sacramento Kings #NBAKings 0
WESTLOSE westlose:20080919 [] New Orleans Hornets #Hornets 0
WESTLOSE westlose:20080920 [] Memphis Grizzlies #Grizzlies 0
WESTLOSE westlose:20080921 [] L.A. Clippers #Clippers 0
WESTWIN westwin:20080922 [] San Antonio Spurs #GoSpursGo 9
WESTWIN westwin:20080923 [] Houston Rockets #Rockets 6
WESTWIN westwin:20080924 [] Memphis Grizzlies #Grizzlies 6
WESTWIN westwin:20080925 [] Golden State Warriors #GSWarriors 6
WESTWIN westwin:20080926 [] Utah Jazz #UtahJazz 5
WESTWIN westwin:20080927 [] L.A. Clippers #Clippers 4
WESTWIN westwin:20080928 [] Phoenix Suns #Suns 3
WESTWIN westwin:20080929 [] Minnesota Timberwolves #TWolves 1
WESTWIN westwin:20080930 [] Dallas Mavericks #Mavs 1
WESTWIN westwin:20080931 [] Oklahoma City #okcthunder 0
WESTWIN westwin:20080932 [] Portland Trailblazers #TrailBlazers 0
WESTWIN westwin:20080933 [] Denver Nuggets #Nuggets 0
WESTWIN westwin:20080934 [] Sacramento Kings #NBAKings 0
WESTWIN westwin:20080935 [] LA Lakers #Lakers 0
WESTWIN westwin:20080936 [] New Orleans Hornets #Hornets 0
west@acc westwinlose> █
```

```
root@acc> user east
password: east
east@acc> table eastwinlose
```

```
east@acc eastwinlose> scan
EASTLOSE eastlose:20080907 [] Boston Celtics #Celtics 3
EASTLOSE eastlose:20080908 [] Atlanta Hawks #Hawks 2
EASTLOSE eastlose:20080909 [] Detroit Pistons #Pistons 1
EASTLOSE eastlose:20080910 [] Miami Heat #MiamiHeat 1
EASTLOSE eastlose:20080911 [] Washington Wizards #Wizards 0
EASTLOSE eastlose:20080912 [] Toronto Raptors #Raptors 0
EASTLOSE eastlose:20080913 [] Indiana Pacers #Pacers 0
EASTLOSE eastlose:20080914 [] Orlando Magic #OrlandoMagic 0
EASTLOSE eastlose:20080915 [] New Jersey Nets #Nets 0
EASTLOSE eastlose:20080916 [] New York Knicks #Knicks 0
EASTLOSE eastlose:20080917 [] Cleveland Cavaliers #Cavs 0
EASTLOSE eastlose:20080918 [] Chicago Bulls #Bulls 0
EASTLOSE eastlose:20080919 [] Milwaukee Bucks #Bucks 0
EASTLOSE eastlose:20080920 [] Charlotte Bobcats #Bobcats 0
EASTLOSE eastlose:20080921 [] Philadelphia 76ers #76ers 0
EASTWIN eastwin:20080922 [] Chicago Bulls #Bulls 18
EASTWIN eastwin:20080923 [] Cleveland Cavaliers #Cavs 15
EASTWIN eastwin:20080924 [] Detroit Pistons #Pistons 7
EASTWIN eastwin:20080925 [] Philadelphia 76ers #76ers 3
EASTWIN eastwin:20080926 [] Washington Wizards #Wizards 2
EASTWIN eastwin:20080927 [] Atlanta Hawks #Hawks 2
EASTWIN eastwin:20080928 [] New York Knicks #Knicks 1
EASTWIN eastwin:20080929 [] Toronto Raptors #Raptors 0
EASTWIN eastwin:20080930 [] Indiana Pacers #Pacers 0
EASTWIN eastwin:20080931 [] Orlando Magic #OrlandoMagic 0
EASTWIN eastwin:20080932 [] New Jersey Nets #Nets 0
EASTWIN eastwin:20080933 [] Miami Heat #MiamiHeat 0
EASTWIN eastwin:20080934 [] Boston Celtics #Celtics 0
EASTWIN eastwin:20080935 [] Milwaukee Bucks #Bucks 0
EASTWIN eastwin:20080936 [] Charlotte Bobcats #Bobcats 0
east@acc eastwinlose> █
```

README

- **User Name: east**
Password: east
- **user Name: west**
password: west
- **East output Table Name: eastwinlose**
- **West output Table Name: westwinlose**

Command used to run the jar:

**bin/tool.sh /home/guest/Documents/accumulo_karthik.jar Main acc guestvb /input
winlose -u root -p acc**